Attorney Docket No. <u>JP919990203</u>

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): Y. Tahara et al.
Docket No.: JP919990203
Serial No.: 09/656,963
Filing Date: September 7, 2000
Group: 2655
Examiner: Michael N. Opsasnick

Title: Methods and Apparatus for Voice Information Registration and Recognized Sentence Specification in Accordance With Speech Recognition

RECEIVED
OCT 0 9 2003
Technology Center 2600

APPEAL BRIEF

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313

Sir:

Applicants (hereinafter referred to as "Appellants") hereby appeal the final rejection of claims 1-4, 6-9 and 11-14 of the above referenced application.

REAL PARTY IN INTEREST

The present application is assigned to International Business Machines Corp., as evidenced by an assignment recorded December 5, 2000 in the U.S. Patent and Trademark Office at Reel 11315, Frame 0734. The assignee, International Business Machines Corp., is the real party in interest.

RELATED APPEALS AND INTERFERENCES

There are no known related appeals and interferences.

STATUS OF CLAIMS

Claims 1-15 are pending in the present application. Claims 5, 10 and 15 are allowable, and claims 1-4, 6-9 and 11-14 stand rejected under 35 U.S.C. §103(a). Claims 1-4, 6-9 and 11-14 are appealed.

STATUS OF AMENDMENTS

There have been no amendments filed subsequent to the rejection.

SUMMARY OF INVENTION

The present invention relates to a word registration method for a speech recognition system and, more particularly, to a method whereby voice is used to specify information displayed on a screen (Specification, page 1, lines 5-8).

By way of example, as recited in claim 1, a voice information registration method, employed by a speech recognition apparatus, comprises the following steps. A sentence group is obtained, which includes a first to an N-th sentence, wherein N is a number equal to or greater than two. A sounds-like spelling is obtained for a word that is included in an i-th sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N. A base form is obtained based on the sounds-like spelling of the word. Finally, the base form is registered in a speech recognition dictionary in correlation with the word.

In an illustrative embodiment, a group of sentences to be recognized is obtained from an application. Using parsing logic, each target sentence to be recognized is divided into words, or speech recognition units. Thereafter, the words in each target sentence are examined to determine whether there are unknown words among them that are not registered in the speech recognition dictionary, but for which the sounds-like spelling is available. If an unknown word is found, a base form, from which the pronunciation is inferred from the sounds-like spelling, is prepared and registered in the speech recognition dictionary. This base form is employed when the voice of a user is recognized who has orally designated one of the sentences (Specification, page 2, lines 4-12)

According to one aspect of the present invention, a voice information registration method is provided, which is employed by a speech recognition apparatus, and with which a voice input device

2

is used (Specification, page 2, lines 13-15). According to another aspect of the present invention, a sentence specification method is provided that is employed by a speech recognition apparatus, and with which a voice input device is used. This sentence specification method has both a registration step and a recognition step (Specification, page 2, lines 24-26).

An unknown word, detected in an extracted sentence, is recognized as a word but is not registered in the speech recognition dictionary. Thus the base form of the unknown word is unknown to the system (Specification, page 17, lines 5-7). A flow diagram showing the unknown word detection processing performed according to an embodiment of the present invention is shown in FIG. 5. A flow diagram showing the processing performed according to an embodiment of the present invention to obtain a base form corresponding to an unknown word and to register the base form in a speech recognition dictionary is shown in FIG. 6. As described in the figures, the present invention allows for the recognition of a sentence that includes words that are not registered in a speech recognition dictionary through the registration process of the present invention (Specification, page 26, lines 10-12).

ISSUES PRESENTED FOR REVIEW

1. Whether claims 1-4, 6-9 and 11-14 are properly rejected under 35 U.S.C. §103(a) as being unpatentable over U.S. Patent No. 5,208,897 to Hutchins (hereinafter "Hutchins"), in view of IBM Technical Disclosure Bulletin, Vol. 35, Issue 1a, p. 59 (hereinafter "TDB").

GROUPING OF CLAIMS

Claims 1-4, 6-9 and 11-14 do not stand or fall together. More particularly, claims 1, 6 and 11 stand or fall together and claims 2-4, 7-9 and 12-14 stand or fall together.

ARGUMENT

Appellants incorporate by reference herein the disclosure of all previous responses filed in the present application, namely, responses dated February 19, 2003 and July 31, 2003.

3

With regard to the issue of whether claims 1-4, 6-9 and 11-14 are properly rejected under 35 U.S.C. §103(a), Appellants respectfully reassert that the cited combination fails to establish a prima facie case of obviousness under 35 U.S.C. §103(a), as specified in M.P.E.P. §2143.

As set forth therein, M.P.E.P. §2143 states that three requirements must be met to establish a prima facie case of obviousness. First, the cited combination must teach or suggest all the claim limitations. Second, there must be some suggestion or motivation to combine reference teachings. Third, there must be a reasonable expectation of success. While it is sufficient to show that a prima facie case of obviousness has not been established by showing that one of the requirements has not been met, Appellants respectfully believe that none of the requirements have been met.

Appellants believe that, given the summary of the invention and the discussion of the deficiencies of the cited combination to follow, it will be evident that the §103(a) rejections should be withdrawn. The invention is directed toward <u>techniques for registering unknown words</u> such that these words may then be used in <u>techniques for recognizing speech uttered by a user</u>.

By way of further explanation, the present specification, at page 1, lines 8-19, describes one problem that the claimed invention addresses:

> As is described in Japanese Unexamined Patent Publication No. Hei 10-320168, the disclosure of which is incorporated by reference herein, a conventional method is available whereby voice is used to specify information displayed on a screen. However, to use this method, a menu or a button in an application, and a sentence in which a link to a web is included must be registered using words that can be recognized by a speech recognition system.
>
> All of the character strings for a menu, in this case, can be statically added to a speech recognition dictionary, but since the web link would tend to be changed daily, coping with such a change would exceed the capabilities of a method for which static registration is employed. In addition, if too many words, more than are necessary, are added to the dictionary, other problems, such as a reduction in the recognition accuracy or an extended processing time, may be encountered.

First, with respect to independent claims 1, 2, 6, 7, 11 and 12, the collective teaching of Hutchins and TDB fails to suggest or render obvious the elements of such claims. For at least this reason, a prima facie case of obviousness has not been established.

4

Hutchins discloses techniques for recognizing speech, and does not disclose registration techniques, as in the claimed invention. More particularly, the abstract of Hutchins discloses a method for speech recognition which includes steps of sampling a speaker's speech and providing speech data sample segments of predetermined length at predetermined sampling intervals based on changes in energy in the speech. TDB discloses a technique for building word models for a speech recognition system that includes a user entering a "sounds-like spelling."

The final Office Action (at paragraph 5) contends that Hutchins discloses the steps/operations of independent claims 1, 2, 6, 7, 11 and 12 including: "obtaining a sentence group . . . sentence;" "obtaining a spelling . . . speech recognition dictionary;" "obtaining a base form . . . word;" and "registering said base form . . . said word." The Office Action again acknowledges that Hutchins does not disclose "sounds-like spellings," however, points out that TDB uses "sounds-like spellings." The Office Action then summarily concludes that it would have been obvious to combine the two references to achieve the claimed invention. Appellants strongly disagree.

The first step of independent claim 1 of the present invention recites "obtaining a sentence group. . . " The Office Action alleges that a section of Hutchins describing word grammar (words to phrases) discloses this step. However, this section of Hutchins discloses the process of combining individual words from digitized speech to form phrases in the word recognition process. This combination of individual words does not disclose the obtaining of groups of sentences in a word registration process.

The second step of independent claim 1 of the present invention recites "obtaining a sounds-like spelling for a word that is included in an i-th sentence, but is not entered in a speech recognition dictionary . . ." The Examiner admits that Hutchins does not disclose the use of a sounds-like spelling. However, the Office Action alleges that a section of Hutchins describing the ASCII spelling of words as output resulting from incoming speech, discloses the remainder of this step. However, ASCII output of Hutchins occurs once the word recognition process is complete and the spoken words were found in a recognition dictionary. Therefore, Hutchins does not disclose providing a spelling of a word that is not entered in a speech recognition dictionary as part of a word registration process.

The third step of independent claim 1 of the present invention recites "obtaining a base form based on said sounds-like spelling of said word . . ." The Office Action alleges that a section of Hutchins that relates to processing subsyllables and syllables discloses this step. However, in the present invention the base form of a word is based on a sounds-like spelling in a word registration process. Further, the base form is from a word not registered in the speech recognition dictionary. This differs significantly from syllable processing in a speech recognition process, as described in Hutchins.

The fourth step of independent claim 1 of the present invention recites "registering said base form in a speech recognition dictionary in correlation with said word." The Office Action again alleges that a section of Hutchins relating to a word recognition process combining subsyllables to form syllables, and combining syllables to form words, discloses this step. These syllables and subsyllables are not being registered as a part of the speech registration process. Thus, Hutchins does not disclose the speech registration elements recited in independent claim 1 or the remaining independent claims of the present invention.

As explained above, Hutchins is directed only toward speech recognition, not unknown word registration. While registered words are eventually used to recognize speech in the present invention, the techniques for realizing the two concepts are significantly different. In response to arguments previously submitted by the Appellants, the final Office Action and Advisory Action contend that Hutchins teaches word recognition and uses the word recognition system to improve the accuracy of the existing vocabulary or to increase the existing vocabulary, thereby inherently teaching word registration. However, Hutchins does not contain the disclosure which is necessary to support a claim rejection on the basis of inherency. According to the Court of Customs and Patent Appeals (CCPA), "Inherency does not mean that a thing might be done, or that it might happen, ...; but it must be disclosed, if inherency is claimed, that the thing will necessarily happen." In re Draeger et al., 150 F.2d 572, 574 (CCPA 1945) (emphasis supplied). Furthermore, the well settled law "requires that inherency may not be established by possibilities and probabilities. The evidence must show that the inherency is necessary and inevitable." Interchemical Corp. v. Watson, 145 F.Supp. 179, 182, 111 USPQ 78, 79 (D. D.C. 1956) (emphasis supplied), aff'd, 251 F.2d 390, 116 USPQ 119 (D.C. Cir. 1958).

Additionally, "in relying upon the theory of inherency, the examiner must provide a basis in fact and/or technical reasoning to reasonably support the determination that the allegedly inherent characteristic *necessarily* flows from the teachings of the applied prior art." Ex parte Levy, 17 USPQ2d 1461, 1464 (Bd. Pat. App. & Inter. 1990) (emphasis in original). As acknowledged by the Examiner, there is an absence of any teaching of word registration in Hutchins. Therefore, there is no reasonable basis for an assertion that a resulting word registration system, having the methods of the present invention, necessarily flows from the system disclosed in the Hutchins reference, and thus is inherent. No such basis and/or technical reasoning has been provided by the Examiner in the Office Action.

Furthermore, with respect to TDB, while discussing the use of "sounds-like spellings," TDB explains that a user may enter a "sounds-like spelling" to assist in building word models. This significantly differs from the steps/operations of the claimed invention since TDB, other than stating that a user may enter a "sounds-like spelling," is silent as how a word model is formed given the "sounds-like spelling." Thus, TDB suggests nothing about obtaining a sentence group, obtaining a sounds-like spelling for a word that is included in an i-th sentence, but is not entered in a speech recognition dictionary, obtaining a base form based on said sounds-like spelling of said word, and registering said base form in a speech recognition dictionary in correlation with said word, as recited in the claimed invention.

Appellants do not assert that they have developed the concept of "sounds-like spellings." Such spellings are known, for example, as evidenced by TDB. However, the automated use of "sounds-like spellings" as recited in the registration techniques of the claimed invention was not known prior to the invention and is clearly not taught or suggested by the combination of Hutchins and TDB.

Also, with specific regard to claims 2, 7 and 12, despite a contention in the Office Action to the contrary, neither Hutchins nor TDB disclose obtaining voice information that is input as a user reads and vocally reproduces a display corresponding to the i-th sentence, as in the claimed invention.

Therefore, as pointed out above, since both Hutchins and TDB fail to teach or suggest the limitations of claims 1, 2, 6, 7, 11 and 12, their combination also fails to do so.

Second, with respect to independent claims 1, 2, 6, 7, 11 and 12, Appellants reassert that no motivation or suggestion exists to combine Hutchins and TDB in a manner proposed by the Examiner, or to modify their teachings to meet the claim limitations. For at least this reason, a prima facie case of obviousness has not been established. In the response to arguments previously set forth by the Appellants, the Examiner states that the motivation for combining the references is so that the user may type in the spelling to improve the recognition process. However, the present invention incorporates sounds-like spellings into a speech registration system, which are used to form base words that are then registered in a recognition dictionary. Therefore, Appellants still fail to see the motivation or suggestion to combine the very specific subsyllable speech recognition techniques of Hutchins with the word model building techniques of TDB. While both references generally relate to aspects of speech recognition, Appellants strongly believe that one ordinarily skilled in the art would not look to the word model building techniques of TDB to find inspiration to improve the very specific subsyllable speech recognition techniques of Hutchins, or vice versa.

Furthermore, the Federal Circuit has stated that when patentability turns on the question of obviousness, the obviousness determination "must be based on objective evidence of record" and that "this precedent has been reinforced in myriad decisions, and cannot be dispensed with." In re Lee, 277 F.3d 1338, 1343 (Fed. Cir. 2002). Moreover, the Federal Circuit has stated that "conclusory statements" by an examiner fail to adequately address the factual question of motivation, which is material to patentability and cannot be resolved "on subjective belief and unknown authority." Id at 1343-1344.

In the final Office Action at paragraphs 5 and 6, the Examiner provides the following statements to prove motivation to combine Hutchins and TDB, with emphasis supplied: "it would have been obvious . . . to modify the teachings of Hutchins with a 'sounds like spelling' technique because it would advantageously allow user to enter the information more accurately that the phonetic pronunciations . . . the motivation is allowing the user to type in the spelling to improve the recognition." Appellants submit that these statements are based on the type of "subjective belief and unknown authority" that the Federal Circuit has indicated provides insufficient support for an obviousness rejection. More specifically, the Examiner fails to identify any objective evidence of record which supports the proposed combination.

Lastly, with respect to independent claims 1, 2, 6, 7, 11 and 12, Appellants reassert that there is no reasonable expectation of success in achieving the present invention through a combination of Hutchins and TDB. For at least this reason, a prima facie case of obviousness has not been established. Despite the assertion in the final Office Action, Appellants do not believe that Hutchins and TDB are combinable since it is not clear how one would combine them. No guidance was provided in the final Office Action as to how the two references can be combined to achieve the present invention. However, even if combined, for the sake of argument, they would not achieve the automated registration techniques of the claimed invention. For example, elements of a word registration system are not described in either reference, and the use of a sounds-like spelling in a word recognition process differs from obtaining a sounds-like spelling in a word registration process for obtaining the base form of a word, as described in the independent claims of the present invention.

Therefore, for at least the reasons given above, Appellants again respectfully request that the §103(a) rejections of independent claims 1, 2, 6, 7, 11 and 12 be withdrawn.

Furthermore, it is respectfully reasserted that the claims which respectively depend from independent claims 2, 7 and 12, i.e., claims 3, 4, 8, 9, 13 and 14, are patentable over the cited combinations. Thus, Appellants again request withdrawal of the §103(a) rejections of said claims.
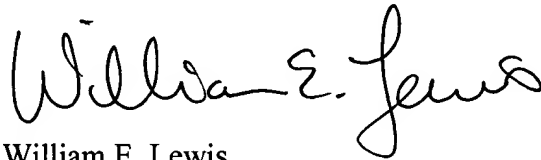
In addition, it is asserted that dependent claims 3, 4, 8, 9, 13 and 14 recite patentable subject matter in their own rights. Claims 3, 8 and 13 recite that the group of sentences is obtained from an application and that a control message corresponding to the i-th sentence is generated and transmitted to the application. There is no application or control message disclosed in Hutchins or TDB. Further, claims 4, 9 and 14 recite that a sounds-like spelling score is stored in correlation with the sounds-like spelling of the word, that a pronunciation score is stored in correlation with the base form, and that the base form is registered in a speech recognition dictionary when a function value that is obtained by using the sounds-like spelling score and the pronunciation score exceeds a threshold value. Again, Hutchins and TDB are completely silent as to the use of any scores stored in correlation with sounds-like spellings and base forms. In response to the arguments previously presented by the Appellants, the Examiner stated that the response fails to comply with 37 C.F.R. §1.111(b) because they amount to a general allegation that the claims define a patentable invention

9

without specifically pointing out how the language of the claims patentably distinguishes them from the references. However, as presented above, the patentable language of the claims is fully presented and not disclosed in the references cited by the Examiner.

In response to arguments previously submitted by the Appellants, an Advisory Action dated August 25, 2003 was issued by the Examiner. However, the Advisory Action does not address all the arguments set forth by the Appellants, for example, the failure of Hutchins to establish a prima facie case of obviousness.

For at least the reasons given above, Appellants respectfully request withdrawal of the §103(a) rejections of claims 1-4, 6-9 and 11-14. Appellants believe that claims 1-4, 6-9 and 11-14 are not obvious in view of Hutchins and TDB. As such, the application is asserted to be in condition for allowance, and favorable action is respectfully solicited.

Respectfully submitted,

Date: October 6, 2003

William E. Lewis
Attorney for Applicant(s)
Reg. No. 39,274
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-2946

APPENDIX

1. A voice information registration method, employed by a speech recognition apparatus, comprising:

(a) obtaining a sentence group, which includes a first to an N-th sentence, wherein N is a number equal to or greater than two;

(b) obtaining a sounds-like spelling for a word that is included in an i-th sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N;

(c) obtaining a base form based on said sounds-like spelling of said word; and

(d) registering said base form in a speech recognition dictionary in correlation with said word.

2. A sentence specification method, employed by a speech recognition apparatus, comprising:

a registration step including:

(a1) obtaining a sentence group, which includes a first to an N-th sentence, wherein N is a number equal to or greater than two,

(a2) obtaining a sounds-like spelling for a word that is included in an i-th sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N,

(a3) obtaining a base form based on said sounds-like spelling of said word, and

(a4) registering said base form in a speech recognition dictionary in correlation with said word, and

a recognition step including:

(b1) obtaining voice information that is input as a user reads and vocally reproduces a display corresponding to said i-th sentence,

(b2) employing said base form to recognize said voice information and to select a speech recognition sentence, and

(b3) comparing said i-th sentence with said selected speech recognition sentence.

11

3. The sentence specification method according to claim 2, wherein said group of sentences is obtained from an application, said method further comprising a step of generating a control message corresponding to said i-th sentence and transmitting said control message to said application.

4. The sentence specification method according to claim 2, wherein a sounds-like spelling score is stored in correlation with the sounds-like spelling of said word, wherein a pronunciation score is stored in correlation with said base form, and wherein, when a function value that is obtained by using said sounds-like spelling score and said pronunciation score exceeds a threshold value, said base form is registered in a speech recognition dictionary.

6. A speech recognition apparatus, comprising:

(a) a sentence specification unit for obtaining a sentence group, which includes a first to an N-th sentence, wherein N is a number equal to or greater than two;

(b) an unknown word detector for obtaining a sounds-like spelling for a word that is included in an i-th sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N;

(c) a base form generator for obtaining a base form based on said sounds-like spelling of said word; and

(d) a speech recognition dictionary to which said base form is stored in correlation with said word.

7. A speech recognition apparatus, comprising:

(a) a sentence specification unit for obtaining a sentence group, which includes a first to an N-th sentence, wherein N is a number equal to or greater than two;

(b) an unknown word detector for obtaining a sounds-like spelling for a word that is included in an i-th sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N;

12

(c) a base form generator for obtaining a base form based on said sounds-like spelling of said word;

(d) a speech recognition dictionary in which said base form is stored in correlation with said word;

(e) a voice input unit for obtaining voice information that is input as a user reads and vocally reproduces a display corresponding to said i-th sentence; and

(f) a speech recognition engine for employing said base form to recognize said voice information and to select a speech recognition sentence;

wherein said sentence specification unit compares said i-th sentence with said selected speech recognition sentence.

8. The speech recognition apparatus according to claim 7, wherein said sentence specification unit obtains said group of sentences from an application, generates a control message corresponding to said i-th sentence, and transmits said control message to said application.

9. The speech recognition apparatus according to claim 7, wherein a sounds-like spelling score is stored in correlation with the sounds-like spelling of said word, wherein a pronunciation score is stored in correlation with said base form, and wherein, when a function value that is obtained by using said sounds-like spelling score and said pronunciation score exceeds a threshold value, said base form is registered in a speech recognition dictionary.

11. A storage medium in which a program for specifying a sentence is stored to be executed by a speech recognition apparatus, said program comprising:

(a) program code for instructing said speech recognition apparatus to obtain a sentence group, which includes a first to an N-th sentence, wherein N is a number equal to or greater than two;

(b) program code for instructing said speech recognition apparatus to obtain a sounds-like spelling for a word that is included in an i-th sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N;

(c) program code for instructing said speech recognition apparatus to obtain a base form based on said sounds-like spelling of said word; and

(d) program code for instructing said speech recognition apparatus to register said base form in a speech recognition dictionary in correlation with said word.

12. A storage medium in which a program for specifying a sentence is stored to be executed by a speech recognition apparatus, said program comprising:

(a) program code for instructing said speech recognition apparatus to obtain a sentence group, which includes a first to an N-th sentence, wherein N is a number equal to or greater than two;

(b) program code for instructing said speech recognition apparatus to obtain a sounds-like spelling for a word that is included in an i-th sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N;

(c) program code for instructing said speech recognition apparatus to obtain a base form based on said sounds-like spelling of said word;

(d) program code for instructing said speech recognition apparatus to register said base form in a speech recognition dictionary in correlation with said word;

(e) program code for instructing said speech recognition apparatus to obtain voice information that is input as a user reads and vocally reproduces a display corresponding to said i-th sentence;

(f) program code for instructing said speech recognition apparatus to employ said base form to recognize said voice information and to select a speech recognition sentence; and

(g) program code for instructing said speech recognition apparatus to compare said i-th sentence with said selected speech recognition sentence.

13. The storage medium according to claim 12, wherein said group of sentences is obtained from an application, and wherein program code is stored to instruct said speech recognition apparatus to generate a control message corresponding to said i-th sentence and to transmit said control message to said application.

14. The storage medium according to claim 12, wherein a sounds-like spelling score is stored in correlation with the sounds-like spelling of said word, wherein a pronunciation score is stored in correction with said base form, and wherein, when a function value that is obtained by using said sounds-like spelling score and said pronunciation score exceeds a threshold value, said based form is registered in a speech recognition dictionary.